

# IOWA STATE UNIVERSITY

## Digital Repository

---

Project AAIMS Technical Reports

Project AAIMS

---

2006

## Reliability, Criterion Validity and Sensitivity to Growth: Extending Work on Two Algebra Progress Monitoring Measures

Serkan Perkmen

*Iowa State University*

Anne Foegen

*Iowa State University, [afoegen@iastate.edu](mailto:afoegen@iastate.edu)*

Jeannette R. Olson

*Iowa State University, [olsonj@iastate.edu](mailto:olsonj@iastate.edu)*

Follow this and additional works at: [http://lib.dr.iastate.edu/edu\\_aaims\\_reports](http://lib.dr.iastate.edu/edu_aaims_reports)



Part of the [Curriculum and Instruction Commons](#), [Higher Education Commons](#), and the [Science and Mathematics Education Commons](#)

---

### Recommended Citation

Perkmen, Serkan; Foegen, Anne; and Olson, Jeannette R., "Reliability, Criterion Validity and Sensitivity to Growth: Extending Work on Two Algebra Progress Monitoring Measures" (2006). *Project AAIMS Technical Reports*. 12.

[http://lib.dr.iastate.edu/edu\\_aaims\\_reports/12](http://lib.dr.iastate.edu/edu_aaims_reports/12)

This Report is brought to you for free and open access by the Project AAIMS at Iowa State University Digital Repository. It has been accepted for inclusion in Project AAIMS Technical Reports by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

---

# Reliability, Criterion Validity and Sensitivity to Growth: Extending Work on Two Algebra Progress Monitoring Measures

## **Abstract**

This report is intended to extend previous work examining the reliability and validity of the Basic Skills and the Content-Analysis-Multiple Choice probes in two Iowa schools districts and to explore the use of the measures for monitoring student progress. One hundred thirty five students in grades nine to twelve participated in the study. Data were gathered from September 2005 to January 2006. Over four months of data collection, students completed two Basic Skills probes and two Content Analysis-Multiple Choice probes each month. We assessed the alternate form reliability and test-retest reliability of both types of probes. Our findings revealed that both types of probes possessed adequate levels of reliability with the Basic Skills probes demonstrating a higher level of reliability than the Content Analysis-Multiple Choice probes. To evaluate the validity of both types of probes, we gathered data from a variety of indicators of students' proficiency in algebra including course grades, teachers' evaluations of student proficiency and growth, and performance on standardized assessment instruments including the Iowa Test of Education Development (ITED) and the Iowa Algebra Aptitude Test (IAAT). We examined both concurrent and predictive validity. Our results revealed moderate correlations between both types of probes and other indicators of students' proficiency in algebra. We examined whether the probes reflected student growth over time. The analyses revealed students grew .51 and .61 points each week on Basic Skills and Content Analysis- Multiple Choice probes, respectively. This result suggests that the Content Analysis-Multiple Choice probes may be slightly more sensitive in evaluating student growth. Finally, we were also interested in exploring whether students' growth on the two types of probes was related to teachers' rating of growth or gain scores on the IAAT. We found a small but significant correlation between teacher ratings of growth and both types of probes. No significant correlations existed between either type of probe and IAAT gain scores.

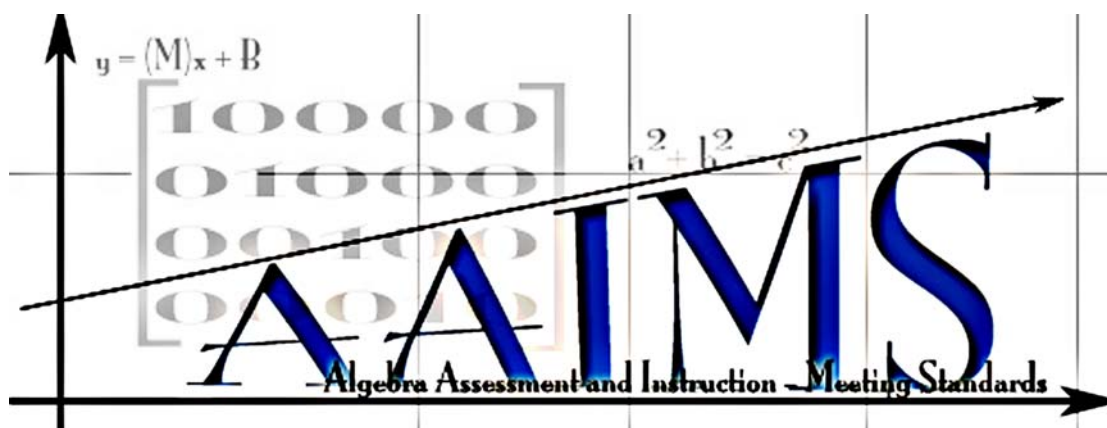
## **Disciplines**

Curriculum and Instruction | Education | Higher Education | Science and Mathematics Education

## **Comments**

Technical Report #12

# PROJECT AAIMS: ALGEBRA ASSESSMENT AND INSTRUCTION – MEETING STANDARDS



Reliability, Criterion Validity and Sensitivity to Growth: Extending Work  
on Two Algebra Progress Monitoring Measures

Technical Report #12

Serkan Perkmen, M.Ed.

Anne Foegen, Ph.D.

Jeannette Olson, M.S.

Iowa State University

September 2006

# PROJECT AAIMS: ALGEBRA ASSESSMENT AND INSTRUCTION – MEETING STANDARDS

## Reliability, Criterion Validity, and Sensitivity to Growth: Extending Work on Two Algebra Progress Monitoring Measures

### Technical Report 12

---

#### Abstract

---

This report is intended to extend previous work examining the reliability and validity of the Basic Skills and the Content-Analysis-Multiple Choice probes in two Iowa schools districts and to explore the use of the measures for monitoring student progress. One hundred thirty five students in grades nine to twelve participated in the study. Data were gathered from September 2005 to January 2006. Over four months of data collection, students completed two Basic Skills probes and two Content Analysis-Multiple Choice probes each month. We assessed the alternate form reliability and test-retest reliability of both types of probes. Our findings revealed that both types of probes possessed adequate levels of reliability with the Basic Skills probes demonstrating a higher level of reliability than the Content Analysis-Multiple Choice probes. To evaluate the validity of both types of probes, we gathered data from a variety of indicators of students' proficiency in algebra including course grades, teachers' evaluations of student proficiency and growth, and performance on standardized assessment instruments including the Iowa Test of Education Development (ITED) and the Iowa Algebra Aptitude Test (IAAT). We examined both concurrent and predictive validity. Our results revealed moderate correlations between both types of probes and other indicators of students' proficiency in algebra. We examined whether the probes reflected student growth over time. The analyses revealed students grew .51 and .61 points each week on Basic Skills and Content Analysis- Multiple Choice probes, respectively. This result suggests that the Content Analysis-Multiple Choice probes may be slightly more sensitive in evaluating student growth. Finally, we were also interested in exploring whether students' growth on the two types of probes was related to teachers' rating of growth or gain scores on the IAAT. We found a small but significant correlation between teacher ratings of growth and both types of probes. No significant correlations existed between either type of probe and IAAT gain scores.

---

## Full Report

---

### Introduction

Previous work in Project AAIMS has established the reliability and criterion validity of three measures for monitoring student progress in algebra. In Technical Report 10, we reported the technical features of the measures when used for static (i.e., one point in time) measurement of student performance. The three measures (Basic Skills, Algebra Foundations, Content Analysis-Multiple Choice) have acceptable levels of reliability and moderate levels of criterion validity. While it is valuable to have measures that can be used at a single point in time, if teachers want to use the measures to track student progress and inform their instructional decisions, it is important that the measures also reflect changes in student performance over time. The study reported here was conducted to replicate technical adequacy findings of the earlier study and to examine the degree to which the measures were sensitive to changes in student performance over time.

### Method

The study described in this report was conducted during the fall semester of the 2005-06 school year in Districts B and C. District B is located in a community of 26,000 people, where the high school currently serves 1,349 students. The majority of students are white (82%), and nearly half are eligible for free and reduced lunch (47%). Eighteen percent of the students are of diverse backgrounds in terms of race, culture and ethnicity. Approximately 15% of the student population is identified as eligible for special education services. District B uses block scheduling, so students complete a traditional course in approximately four and one half months. Each instructional period is approximately 90 minutes in length, and the school day consists of four instructional periods.

District C is located in a predominantly rural area and serves approximately 17,700 residents in five small towns and a Native American Settlement community. The high school enrolls 488 students in grades 9 through 12. Thirty-nine percent of the districts' students are of diverse backgrounds in terms of race, culture and ethnicity. Approximately 45% of the student population is eligible for free and reduced lunch. Approximately 15% of the student population has been identified as students eligible for special education services. Like District B, District C also uses block scheduling with a 90 minute period and four instructional periods in each school day.

Data for the study were gathered from September 2005 to January 2006. During the first data collection session, students completed the algebra criterion measure. All data collection activities involving students were completed during regular class time. Teachers administered all algebra probes.

### Participants

One hundred two students in District B and 33 students in District C participated in the study. Written parental/guardian consent and written student assent were obtained for all of these students using procedures approved by Iowa State University's Human Subjects Review Committee. Descriptions of the participating students from each district are provided in Tables 1 and 2.

Table 1. Demographic Characteristics of Student Participants by Grade Level for District B

	Total	Grade 9	Grade 10	Grade 11	Grade 12
N	102 <sup>1</sup>	62	29	8	2
Gender					
Male	63	36	18	6	2
Female	39	26	11	2	0
Ethnicity <sup>2</sup>					
White	83	50	23	8	1
Black	10	7	3	0	0
Hispanic	6	4	1	0	1
Native Am.	0	0	0	0	0
Asian	2	0	2	0	0
Lunch					
Free/Red	38	29	8	1	0
Disability					
IEP	15	7	6	2	0

<sup>1</sup>One participant's grade level information was not available.

<sup>2</sup>We did not obtain one participant's ethnicity information.

Table 2. Demographic Characteristics of Student Participants by Grade Level for District C

	Total	Grade 9	Grade 10	Grade 11	Grade 12
N	33	22	5	4	2
Gender					
Male	15	11	3	1	0
Female	18	11	2	3	2
Ethnicity					
White	25	16	5	3	1
Black	0	0	0	0	0
Hispanic	3	2	0	0	1
Native Am.	5	4	0	1	0
Lunch					
Free/Red	15	11	0	3	1
Disability					
IEP	3	1	1	1	0

As the data in Tables 1 and 2 indicate, many of the participants (an average of 80%) were white and an average of 62% were in ninth grade, the traditional grade in which students in these districts complete algebra. Thirty-seven and 45 percent participated in federal free or reduced lunch programs in Districts B and C, respectively and 15% and 9% of the participating students in Districts B and C, respectively, were students with disabilities who were receiving special education services. In District B, 87 students were enrolled in Algebra 1A and 15 in Algebra 1B. Algebra 1A and 1B is an option available in District B in which students complete half the content of a traditional Algebra 1 course in a single course. A similar arrangement was offered

in District C in which a small number of students were enrolled in an algebra course that was only half the duration of a typical block class (45 minutes) and lasted for the entire school year. For simplicity, we use the same Algebra 1A/1B terms from District B to refer to this option in both districts. In District C, 18 students were enrolled in Algebra 1 and 15 students were in Algebra 1A. Due to the small number of students in District C participating in the study, data from students in the two schools were combined for statistical analysis purposes.

Additional Information on Students with Disabilities. Because exploring the applicability of the algebra probes to students with disabilities is an important part of Project AAIMS, additional information about the 18 students with disabilities participating in the project is provided in Table 3.

Table 3. Descriptive Information on the Programs of Students with Disabilities

Characteristic	Quantification
Disability category	14 Entitled Individual (EI) 3 Learning Disability 1 Deaf/Hard of Hearing, Severe/Profound
% time in general education	Range = 45 –100%; Mean = 93.72%
# of students with math goals	5
# of students receiving math instruction in general education classes	18

In algebra, students with disabilities earned mean grades of 1.66 [C-] (range 0.00 [F] to 4.00 [A]). In Districts B and C, the Iowa Tests of Educational Development are used as a district-wide assessment. On average, students with disabilities obtained national percentile rank scores of 37 and 41 in Concepts/Problem Solving and Computation, respectively.

### Measures

Algebra Progress Monitoring Measures. Two algebra measures were examined in this study; sample copies of each are provided in the Appendix. The following paragraphs summarize the characteristics of each of the two types of algebra measures used in the study.

#### *Probe A: Basic Skills Measure*

Probe A is designed to assess the ‘tool skills’ or basic skills that students need to be proficient in algebra. Just as elementary students’ proficiency with basic facts is associated with their ease in solving more complex problems, we hypothesize that there are some basic skills in algebra that serve as indicators of overall proficiency. In our discussions with teachers, they frequently commented that many students had difficulty with integers and with applying the distributive property. The items included in the *Basic Skills* measure address solving simple equations, applying the distributive property, working with integers, combining like terms and applying proportional reasoning. The *Basic Skills* probe includes many skills one would assume that students proficient in algebra would be able to complete with reasonable levels of automaticity. We have created six parallel forms of this probe. Students have five minutes to work on this probe. Each *Basic Skills* probe consists of 60 items; each item is scored as one point if it is answered correctly.

### *Probe E: Content Analysis-Multiple Choice Measure*

The *Content Analysis-Multiple Choice* measure is a variation of an earlier measure that used a constructed response format (see Technical Report 7 for additional details on this measure). Our rationale for including a multiple choice option is to see if this format improves scoring efficiency (and potentially interscorer agreement). Another goal is to familiarize students with the multiple-choice format used on district-administered assessments. Like the constructed-response version, this probe consists of 16 items that correspond to different chapters in the textbook that is used in all three districts. This probe includes problems from chapters 1 to 8, rather than from the entire textbook. There are six parallel forms for this probe. Students have seven minutes to work on the Content Analysis-Multiple Choice probes.

Scoring for the *Content Analysis-Multiple Choice* probes is done by comparing student responses to a rubric-based key created by the research staff. Each of the 16 problems is worth up to three points. Students earn full credit (three points) by circling the correct answer from among the four alternatives. If students circle an incorrect response and do not show any work, their answer is considered a 'guess;' the total number of guesses is recorded for each probe. In cases where students show work, the scorer compares the student's work to the rubric-based key, and determines whether the student has earned 0, 1, or 2 points of partial credit. The number of points earned across all 16 problems and the number of guesses are recorded and entered in the data files. A final score is computed by subtracting the number of guesses from the total number of points earned on the probe.

Criterion Measures. In order to evaluate the criterion validity of the algebra progress monitoring measures, we gathered data on a variety of other indicators of students' proficiency in algebra. Some of these measures were based on students' performance in class and their teachers' evaluation of their proficiency. Other measures reflected students' performance on standardized assessment instruments.

The classroom-based measures included grade-based measures and teacher ratings. Each student's *algebra grade*, the grade s/he earned in algebra during the fall semester of the 2005-2006 school year, was recorded using a four-point scale (i.e., A = 4.0, B = 3.0). In District B, students earned a course grade at the end of the fall semester because they had completed a full course in that semester due to the block scheduling format. In District C, students earned grades at four points during the school year (two of which fell during the midpoint and end of the present study). For District C students, we averaged the grades they earned during the duration of the study and used this average as their course grade.

We also included the teachers' evaluations of student proficiency in algebra by asking each teacher to complete a *teacher rating of proficiency* for all the students to whom she/he taught algebra. These ratings were completed at the beginning of the course. Student names were alphabetized across classes to minimize any biases that might be related to particular class sections. Teachers used a 5-point Likert scale (1=low proficiency, 5= high proficiency) to rate each student's proficiency in algebra in comparison to same-grade peers. This enabled us to see if there was a relationship between growth that students showed on both types of probes and the teachers' evaluations of student growth.

Student performance on standardized, norm-referenced assessments was evaluated using school records and with an algebra instrument administered as part of the project. In Districts B and C, students complete the *Iowa Tests of Educational Development* (ITED) each year. District records were used to access students' scores on these instruments; national percentile ranks were



used for the analyses. We recorded the Concepts/Problems subtest score (which was identical to the Math Total score) and the Computation subtest score.

Because the district-administered measure did not provide a direct assessment of algebra, so we also administered the *Iowa Algebra Aptitude Test* (IAAT) and recorded both raw scores and national percentile rank scores. This norm-referenced instrument is typically used to evaluate the potential of 7<sup>th</sup> grade students for successful study of algebra in 8<sup>th</sup> grade. Although we recognized the limitations of using this aptitude measure, we were unable to identify a norm-referenced test of algebra achievement. We had some concerns that there might be ceiling effects when using this measure, but these concerns proved to be unwarranted. We used the percentile rank scores in our concurrent and predictive validity analyses to parallel the analyses involving the ITED,

Growth Measures. One of the major goals of the AAIMS project is to determine the extent to which the two types of measures reflect student growth over time. We were also interested in exploring whether the growth that students showed on the probes was associated with other indicators of growth. To accomplish these goals, we gathered data using three types of measures reflecting students' growth: *probe slope*, *teacher rating of growth*, and *IAAT gain score*.

The first type of growth measure, which we called *probe slope* reflects the growth that students showed on both types of probes over the semester. We used ordinary least square regression to calculate each student's slope on each measure. The obtained slope values were calculated to reflect the amount of weekly progress a student demonstrated on a probe type. The second type of measure was the *teacher rating of growth*. At the end of the semester, we asked teachers rate all the students in their algebra classes. Student names were alphabetized across class periods to minimize any biases that might be related to particular sections. Teachers used a 5-point Likert scale to rate each student's growth in algebra in comparison to same-grade peers. A rating of 1 indicated minimal or no growth, while a rating of 5 represented unusually high growth in comparison to peers. The third type of measure was the *IAAT gain score*, which was calculated by subtracting the total scale raw score for the IAAT form taken at the end of the semester from the total scale raw score for the IAAT form taken at the beginning of the semester. All students in the project completed Form A of the IAAT at the beginning of the study and Form B at the end. Using correlational analysis, we examined the relationships among these growth variables.

### Procedures

Project AAIMS research staff visited each class at the beginning of the school year to present information about the study and gather informed consent. Students completed student assent forms during class and were given parent consent forms to take home. Teachers offered extra credit to students for returning signed consent forms (regardless of whether parents provided or withheld consent). The research staff also administered the Iowa Algebra Aptitude Test at the beginning and end of the study. Teacher ratings forms were distributed at the beginning (initial teacher rating of student proficiency) and the end (teacher rating of growth) of the study and collected by project staff.

The algebra probes were administered during a portion of each class period. Because Districts B and C use block scheduling, each period was approximately 90 minutes in length. Teachers administered probes according to a schedule with one-week intervals during which they

were to give two forms of one type of probe. Some teachers opted to give both probes on the same day; other teachers gave the two probes on two different days. Some of the teachers were unable to administer probes as scheduled for the entire semester. The Xs shown in Table 4 indicate that the teacher administered the probe. As one can see from the table, Teacher 4 was unable to administer the last two probes at the end of the semester. Teacher 5 was unable to administer probes in December and January.

Table 4. Administration Schedule for Probe Forms by Period

Time/Period	Probe	District B				District C	
		T 1	T 2	T 3	T 4	T 5	T 6
Mid-Sep	A-31	X	X	X	X	X	X
Mid-Sep	A-32	X	X	X	X	X	X
End-Sep	E-31	X	X	X	X	X	X
End-Sep	E-32	X	X	X	X	X	X
Mid-Oct	A-33	X	X	X	X	X	X
Mid-Oct	A-34	X	X	X	X	X	X
End-Oct	E-33	X	X	X	X	X	X
End-Oct	E-34	X	X	X	X	X	X
Mid-Nov	A-35	X	X	X	X	X	X
Mid-Nov	A-36	X	X	X	X	X	X
End-Nov	E-35	X	X	X	X	X	X
End-Nov	E-36	X	X	X	X	X	X
Mid-Dec	A-31	X	X	X	X		X
Mid-Dec	A-32	X	X	X	X		X
Mid-Jan	E-31	X	X	X			X
Mid-Jan	E-32	X	X	X			X

Note. A-31 to A-36 denote *Basic Skills* probes; E-31, to E-36 denote *Content Analysis-Multiple Choice* probes

### Scoring Reliability

We hired and trained four pre-service teachers (subsequently referred to as “scorers”) to score the probes. The hiring process included a demonstration of correct scoring procedures for each type of probe and guided practice activities in which scorers worked with actual student papers. A final activity was the independent scoring of 10 student papers for each of the probe types. We used these probes to evaluate scoring reliability. For each probe, an answer-by-answer comparison was conducted and an interscorer reliability estimate was calculated by dividing the number of agreements by the total number of answers scored. These individual probe agreement percentages were then averaged across all the selected probes of a common type to determine an overall average. After training, the scorers’ mean interscorer agreement rates were 98.95% for the Basic Skills probes (range = 98.45% to 99.63%) and 95.84% for the Content Analysis-Multiple Choice probes (range = 94.27% to 96.90%). Scorers were informed that we would be checking their scoring accuracy levels throughout the project; they were able to earn bonus pay for maintaining high levels (i.e., >96% agreement) of accuracy in their scoring.

Following training, each scorer was assigned five classes with two forms of a probe per class to score (a total of 10 class sets of probes twice each month). Readers should note that the total of 20 classes includes additional algebra classes in Project AAIMS whose data are reported

in other technical reports. Scorers also completed the data entry for the classes they were scoring. For each scorer, we conducted a scoring reliability on two of the ten class sets in each scoring period (i.e., twice each month) by re-scoring all of the probes in those sets. The results of these interscorer reliability analyses are reported in the following section.

## **Results**

### Scoring Reliability

Interscorer agreement rates revealed that scorers had high reliability on both types of probes. A total of 112 interscorer reliability checks were conducted across the four scorers throughout the 2005-06 school year. The range of agreement for Basic Skills probes was between 98.1% and 100% with a mean of 99.1%. For Content Analysis-Multiple Choice probes, the interscorer agreement rates ranged from 94.5% to 100%, with a mean of 98.7%.

### Descriptive Data on Score Ranges and Distributions

Table 5 lists the ranges, means, and standard deviations for each of the probes. For the *Basic Skills* probes, the number of correct answers was recorded. The total number of points possible for this probe was 60. On the *Content Analysis-Multiple Choice* probes, the correct score represents the number of points earned on the probe (each of the 16 problems was worth up to 3 points) and the guess score represents the number of guess responses. The total possible correct and guess scores were 48 and 16, respectively.

A close examination of Table 5 reveals two important points. First, both scores and correct responses on the Basic Skills and Content Analysis-Multiple Choice probes gradually increased as the semester went on. This finding suggests that students were improving on their proficiency in completing the types of problems on each of the probes. Second, the standard deviations were substantial (one-third to one-half the magnitude of the means), suggesting that the measures would be beneficial in distributing students based on scores obtained on both probes. This finding is especially important if the probe data are to be used to identify students who are especially strong or weak in algebra.

We also examined whether scores obtained on the Basic Skills and Content Analysis-Multiple Choice probes differed by class type. As discussed earlier, students in three types of classes were participating in the study. Students in Algebra 1 were enrolled in a typical algebra course that was completed in a single semester (daily 90 minute periods on a block schedule). Students in Algebra 1A were completing the first half of Algebra 1 during the period of this study, while students in Algebra 1B were completing the second half of Algebra 1 during the study. We hypothesized that students in Algebra 1B would have the highest initial levels of performance (as they had already completed the first half of Algebra 1), but that the performance levels of Algebra 1 students would rise to similar levels by the end of the study. The means and standard deviations by class type are reported in Table 6 (Basic Skills) and Table 7 (Content Analysis-Multiple Choice).

Table 5. Descriptive Data for Algebra Probes Across Administration Sessions – Raw Scores

<b>Time Period</b>	<b>Probe</b>	<b>N</b>	<b>Score</b>	<b>Range</b>	<b>Mean</b>	<b>Standard Deviation</b>
<i>Algebra Basic Skills Probes</i>						
Mid-September	A-31	125	5 min.	1 - 41	11.06	5.68
	A-32	116	5 min.	3 - 41	12.21	6
Mid-October	A-33	129	5 min.	3 - 38	12.68	6.63
	A-34	126	5 min.	3 - 54	14.20	7.01
Mid-November	A-35	102	5 min.	4 - 54	14.76	7.72
	A-36	103	5 min.	3 - 58	16.74	8.30
Mid-December	A-31	87	5 min.	2 - 59	18.34	8.90
	A-32	87	5 min.	3 - 57	18.62	9.33
<i>Content Analysis-Multiple Choice</i>						
End-September	E-31	125	Correct	3 - 36	13.60	7.05
	E-31	125	Guess	0 - 13	2.16	2.50
	E-32	126	Correct	3 - 39	16.84	8.12
	E-32	126	Guess	0 - 14	2.48	2.69
End-October	E-33	106	Correct	0 - 41	17.83	7.89
	E-33	106	Guess	0 - 15	1.90	2.65
	E-34	105	Correct	3 - 40	16.33	7.69
	E-34	105	Guess	0 - 13	2.47	2.66
End-November	E-35	97	Correct	4 - 42	18.03	8.91
	E-35	97	Guess	0 - 13	2.27	2.65
	E-36	96	Correct	3 - 43	20.42	8.41
	E-36	96	Guess	0 - 10	2.10	2.40
Mid-January	E-31	74	Correct	3 - 42	23.74	8.40
	E-31	74	Guess	0 - 13	1.66	2.24
	E-32	77	Correct	0 - 48	25.90	9.14
	E-32	77	Guess	0 - 13	1.90	2.23

Table 6. Descriptive Data for Basic Skills Probes by Class Type

<b>Time Period/Class Type</b>	<b>N</b>	<b>Range</b>	<b>Mean</b>	<b>Standard Deviation</b>
Mid-September				
<i>Algebra 1A</i>	101	2 - 24	10.43	4.28
<i>Algebra 1B</i>	13	3 - 41	16.00	9.51
<i>Algebra 1</i>	18	5 - 28	14.39	5.80
Mid-October				
<i>Algebra 1A</i>	97	4 - 33	12.64	5.65
<i>Algebra 1B</i>	15	4 - 46	18.20	10.34
<i>Algebra 1</i>	18	3 - 23	12.72	5.72
Mid-November				
<i>Algebra 1A</i>	78	3.5 - 37	14.89	6.78
<i>Algebra 1B</i>	12	6 - 56	21.79	13.11
<i>Algebra 1</i>	14	8.5 - 24	15.21	4.96
Mid-December				
<i>Algebra 1A</i>	70	3 - 47	16.88	7.36
<i>Algebra 1B</i>	12	9 - 58	26.71	13.08
<i>Algebra 1</i>	10	8 - 31	18.15	7.23

Table 7. Descriptive Data for Content Analysis-Multiple Choice Probes by Class Type

<b>Time Period/Class Type</b>	<b>N</b>	<b>Range</b>	<b>Mean</b>	<b>Standard Deviation</b>
End-September				
<i>Algebra 1A</i>	95	0 – 23	10.15	4.90
<i>Algebra 1B</i>	15	10 – 34	21.43	8.16
<i>Algebra 1</i>	17	9 - 35.5	20.44	7.31
End-October				
<i>Algebra 1A</i>	81	0 – 27	13.31	6.32
<i>Algebra 1B</i>	12	10 – 36	21.58	9.25
<i>Algebra 1</i>	15	8 – 34	19.27	6.68
End-November				
<i>Algebra 1A</i>	78	2 – 31	14.11	6.79
<i>Algebra 1B</i>	12	13 – 38	26.13	8.46
<i>Algebra 1</i>	12	11.5 – 42.5	27.21	8.90
Mid-January				
<i>Algebra 1A</i>	57	0 – 40	20.68	8.14
<i>Algebra 1B</i>	12	9 – 41	27.46	10.14
<i>Algebra 1</i>	10	23 – 44	29.20	5.42

The data in Table 6 reveal increases from month to month for students in Algebra 1A and 1B on both probes. Students in Algebra 1, however, had means scores that declined from September to October, but then increased in each subsequent month. On the Basic Skills probes, mean scores for Algebra 1B students were higher than those of the other types of classes for each

time period. This finding was not consistent with our initial hypothesis. The Content Analysis-Multiple Choice data in Table 7 are consistent with our hypothesis, as the Algebra 1B students obtained a higher mean than Algebra 1 students in September and October, but the scores were quite similar in November, with the Algebra 1 students obtained a higher mean at the end of the study.

### Reliability of Probe Scores

The alternate form reliability of individual probes was evaluated by examining the correlation between two forms of a probe given during the same data collection session. We hypothesized that as the semester went on and students became more familiar with the probes, alternate form reliabilities would increase. In Table 8 we present the results of the alternate form reliability analyses. The data reveal that the Basic Skills probes possessed higher reliabilities than the Content Analysis-Multiple Choice probes. As we predicted, the alternate form reliabilities generally increased as the semester went on.

Table 8. Alternate Form Reliability Results for Single Probes

Time Period	Probes	Reliability
<i>Basic Skills</i>		
Mid-September	A-31 and A-32	.83
Mid-October	A-33 and A-34	.82
Mid-November	A-35 and A-36	.89
Mid-December	A-31 and A-32	.91
<i>Content Analysis-Multiple Choice</i>		
End-September	E-31 and E-32	.66
End-October	E-33 and E-34	.61
End-November	E-35 and E-36	.85
Mid-January	E-31 and E-32	.77

Note: All correlations were significant at  $p < .01$ .

We assessed the test retest reliability of the probes by examining the correlation between the mean of two forms of a probe administered across two data collection time periods. For example, the two scores on the Basic Skills probes administered in mid-September were averaged and then correlated with the mean of the two scores on the Basic Skills probes administered in mid-October. Table 9 presents the results of the test-retest reliability analyses. In general, we found that the Basic Skills probes possessed higher test-retest reliabilities than the Content Analysis-Multiple Choice probes. We also hypothesized that as the semester went on, test-retest reliability would increase. Our findings in Table 9 only supported this hypothesis for Basic Skills probes, not for Content Analysis-Multiple Choice probes. The test-retest reliability of the Content Analysis-Multiple Choice probes was relatively constant across the study.

While these coefficients are lower than we would like, particularly for the Content Analysis-Multiple Choice measure, it is important to note that the test-retest period for these scores spanned four to six weeks. This duration is much longer than the five to seven day period often used to evaluate test-retest reliability. Our previous work with these measures using more typical test-retest time frames (see Technical Report 10) produced reliability estimates of .85 and .80 for the Basic Skills and Content Analysis-Multiple Choice measures, respectively.

Table 9: Test-Retest Reliability Results for Aggregated Probes

Time Period	Reliability
<i>Basic Skills</i>	
Mid-September and Mid-October	.75
Mid-October and Mid-November	.83
Mid-November and Mid-December	.89
<i>Content Analysis-Multiple Choice</i>	
End-September and End-October	.75
End-October and End-November	.77
End-November and Mid-January	.75

Note: All correlations were significant at  $p < .01$ .

### Concurrent Validity

The concurrent validity of the measures was examined by correlating scores on the probes with the criterion measures that served as additional indicators of students' proficiency in algebra. The indicators we used included teachers' evaluations of student proficiency and scores obtained from a norm-referenced test of algebra aptitude, the Iowa Algebra Aptitude Test (IAAT). We correlated students' scores on the September probes with the fall teacher rating of student proficiency and the IAAT administered at the beginning of the study (labeled "Pre-IAAT" below). We also correlated students' scores at the end of the study (mid-December and mid-January) with their scores on the IAAT administered at the end of the study ("Post-IAAT"). Correlation coefficients ranging from .44 to .60 (see Table 10) revealed that both scores on the Basic Skills and the Content Analysis-Multiple Choice probes were moderately correlated with criterion measures, which supported their concurrent validity. Student performance on the probes was more strongly correlated with their scores on the IAAT than with their teachers' initial judgments of their proficiency in algebra at the beginning of the course. In addition, the coefficients with the IAAT increased from the beginning to the end of the course, suggesting that as students became more proficient in algebra (and likely obtained higher scores on the probes), their performance was more closely matched to their performance on the IAAT.

Table 10. Concurrent Validity Results for All Students

Time Period	Fall Teacher Rating	Pre-IAAT	Post-IAAT
<i>Basic Skills</i>			
Mid-September	.47	.54	
Mid-December			.60
<i>Content Analysis-Multiple Choice</i>			
End-September	.44	.57	
Mid-January			.58

Note: All of the correlations were significant at .01

We also examined concurrent validity to see if the magnitude of the correlation coefficients differed by class type. These results are reported in Table 11. The majority of the coefficients were in the low moderate to moderate range (.46 to .84). Readers are cautioned to

interpret these results with caution given the small sample sizes ( $N < 20$ ) for the Algebra 1B and Algebra 1 groups. For the Basic Skills probes, we found some minor differences by class type in comparison to the overall coefficients. For the September Basic Skills measures, we found that students in Algebra 1B demonstrated a stronger relation with fall teacher ratings than reflected in the overall average, while Algebra 1 students demonstrated a weaker than average relation. For the pre-IAAT, Algebra 1B and Algebra 1 students both demonstrated stronger relations than the overall average, while those in Algebra 1A demonstrated weaker relations. For the December Basic Skills probes, all groups obtained correlations coefficients that were smaller than those obtained for the overall group.

Table 11. Concurrent Validity by Class Type

Time Period /Class Type	Fall Teacher Rating			Pre-IAAT			Post-IAAT		
	N	Corr	P	N	Corr	p	n	Corr	p
<i>Basic Skills</i>									
Mid-September									
Algebra 1A	83	.47	< .01	96	.35	< .01			
Algebra 1B	13	.57	.04	12	.67	.02			
Algebra 1	18	.35	.15	18	.62	< .01			
Mid-December									
Algebra 1A							67	.59	< .01
Algebra 1B							12	.46	.13
Algebra 1							6	.43	.40
<i>Content Analysis-Multiple Choice</i>									
End-September									
Algebra 1A	82	.37	< .01	91	.29	< .01			
Algebra 1B	14	.65	.01	13	.84	< .01			
Algebra 1	17	.71	< .01	17	.18	.49			
Mid-January									
Algebra 1A							54	.53	< .01
Algebra 1B							12	.62	.03
Algebra 1							7	.30	.51

For the Content Analysis-Multiple Choice probes, the September data were most strongly correlated for the Algebra 1B students (those who had completed the first half of Algebra 1 prior to the start of the study). The relations between the Content Analysis-Multiple Choice scores and the pre-IAAT were particularly low for students who had not previously completed any algebra coursework (Algebra 1A and Algebra 1). The December Content Analysis-Multiple Choice scores were less related to post-IAAT scores for students in Algebra 1 than for students in the other course options.

### Predictive Validity

We examined the predictive validity of the measures by correlating scores on the probes with the criterion measures that served as additional indicators of students' proficiency in algebra. The indicators we used included Post IAAT scores, teacher rating of growth, algebra grade and ITED scores. The results, reported in Table 12, indicate that both probes were



significantly correlated with all of the criterion measures. In general, predictive validity correlation coefficients were in the low to moderate range (.25 to .59). Both types of probes predicted post IAAT scores better than the other indicators. Readers should note that the ITED includes extremely limited content in algebra; as such, the low levels of correlation between algebra probe scores and ITED scores did not surprise us.

Table 12. Predictive Validity Results for All Students

Time Period	Post-IAAT	Teacher Growth Rating	Algebra Grade	ITED Comp	ITED Concept
<i>Basic Skills</i>					
Mid-September	.59	.34	.33	.33	.36
<i>Content Analysis-Multiple Choice</i>					
End-September	.57	.37	.41	.32	.25

*Note: All the correlations were significant at .01*

We also examined predictive validity to see if the correlation coefficients differed by class type. These results are presented in Table 13. As with the earlier analyses by class type, readers are reminded to interpret the results with caution given the small sample sizes for Algebra 1B and Algebra 1. For the Basic Skills probes, scores were more strongly related to the Post-IAAT than to the other measures for the Algebra 1A and Algebra 1 classes. For the Algebra 1B classes, the strongest relation was found with the teacher growth rating. We found that the Content Analysis-Multiple Choice probes demonstrated very high predictive validity for Algebra 1B students by showing a high correlation with teacher rating of growth ( $r = .68$ ), students' algebra grade ( $r = 0.86$ ), IAAT ( $r = .82$ ) and ITED comp ( $r = .71$ ). Student scores on the Content Analysis-Multiple Choice probes were most strongly related to the Post-IAAT for the Algebra 1A students ( $r = .42$ ) and to algebra grades for the Algebra 1 students ( $r = .68$ ).

Table 13. Predictive Validity by Class Type

Time Period /Class Type	Post-IAAT			Teacher Growth Rating			Algebra Grade			ITED Comp			ITED Concept		
<i>Basic Skills</i>	N	<i>r</i>	<i>p</i>	n	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>	n	<i>r</i>	<i>p</i>
Mid-September															
<i>Algebra 1A</i>	78	.56	< .01	86	.26	.01	99	.31	< .01	100	.39	< .01	100	.43	< .01
<i>Algebra 1B</i>	11	.44	.18	10	.50	.14	13	.47	.11	13	.49	.17	13	.23	.46
<i>Algebra 1</i>	8	.65	.08	18	.36	.15	18	.30	.23	13	.10	.73	13	.35	.24
<i>Content Analysis-Multiple Choice</i>															
End-September															
<i>Algebra 1A</i>	75	.42	< .01	82	.26	.02	94	.39	< .01	94	.40	< .01	94	.28	< .01
<i>Algebra 1B</i>	12	.82	< .01	11	.68	.02	15	.86	< .01	15	.71	< .01	15	.18	.51
<i>Algebra 1</i>	8	.28	.50	17	.49	.05	17	.68	< .01	13	.03	.90	13	.23	.46

## Growth

As we considered our hypotheses regarding students' growth on the measures, we assumed that type of class would play an important role. Algebra 1B begin their course having already studied the content of the first half of Algebra 1. We assumed that these students would demonstrate the highest initial levels of performance, but that their rate of growth might not be as high as students in Algebra 1. Given our experiences in the Project AAIMS schools, students who enroll in the Algebra 1A/1B option typically prefer a slower paced course and are not as strong in mathematics as students who enroll in Algebra 1. Given this assumption, we expected that while Algebra 1 students might have lower initial levels of performance on the measures, their rates of growth would likely be steeper. Finally, we anticipated that students in Algebra 1A would have lower initial performance rates than either of the other two class types and would demonstrate a rate of growth similar to Algebra 1B students. To explore these hypotheses further, we conducted three types of analyses of student growth.

In our initial examination of student growth, we plotted the mean score for students in each type of class every month (i.e., the data presented in Table 6) to see if the probes reflected growth similarly. Figure 1 displays the Basic Skills means for each probe administration for each of the three types of classes.

Figure 1. Monthly Growth Shown in Basic Skills Probes

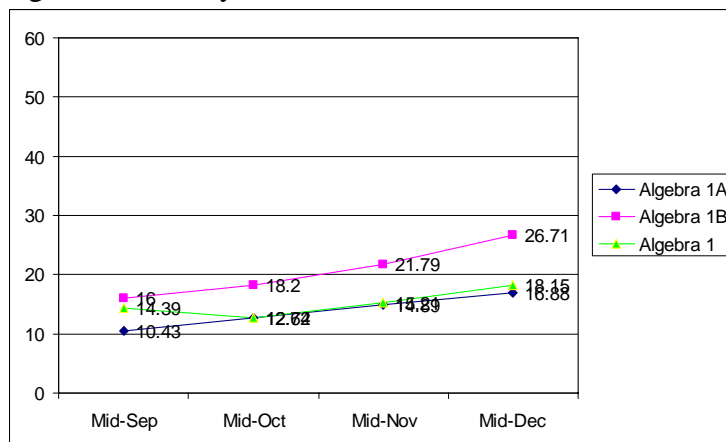
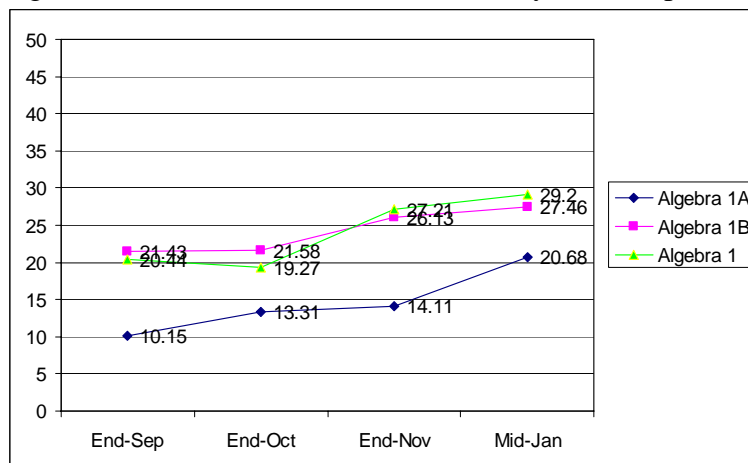


Figure 1 reveals a similar pattern of consistent increases for the Algebra 1 and Algebra 1B groups. In contrast, the Algebra 1A means showed an initial decline, followed by small increases. We next created a similar figure (Figure 2) displaying the results for students in each class type on the Content Analysis-Multiple Choice probes.

Figure 2. Growth Shown in Content Analysis-Multiple Choice Probes



Unlike our findings for the Basic Skills probes, the Content Analysis-Multiple Choice probe data revealed an apparent difference among class types in terms of showing growth. For Algebra 1A, the mean score in the last administration (Mid January Mean = 20.68) was almost twice as high as the mean score in the first administration (End-September Mean = 10.15). For the other two types of classes this ratio was about 1.5. However, if we do not take into account scores in the last administration, there is no seemingly difference among class types.

Our second set of analyses relied on individual students' slope data, rather than the group means. As described earlier, students' weekly growth was calculated using ordinary least square regression. We examined slope values for all students (any students who had data for probes on at least two occasions) and "non-drop" students separately. Non-drop students are those who did not drop their algebra class prior to the end of the study. We were concerned that students who dropped the class might generate slopes that were calculated using just two data points and could potentially be misleading. Table 14 presents the range, mean, and standard deviation of slopes for the Basic Skills and Content Analysis-Multiple Choice probes for all students and for non-drop students.

Table 14. Descriptive Statistics for Weekly Slope Values on Both Probes

	All Students				Non-Drop Students			
	N	Range	Mean	SD	N	Range	Mean	SD
Basic Skills	131	-2.3 – 3	.46	.62	105	-.8 – 2.1	.51	.48
Content Analysis-Multiple Choice	112	-1.3 – 2.75	.63	.68	104	-1.31 – 2.26	.61	.64

The data in Table 14 reveal that removing students who dropped the class resulted in a small increase in mean scores for the Basic Skills probes and a very slight decrease for the Content Analysis-Multiple Choice probes. It is important to note that the sample size changes more dramatically for Basic Skills than for Content Analysis-Multiple Choice probes when the

analyses are limited to non-drop students. Table 14 also indicates that non-drop students grow .51 and .61 points each week on Basic Skills and Content Analysis- Multiple Choice probes, respectively. This result suggests that Content Analysis-Multiple Choice probes may be slightly more sensitive in detecting student growth. This difference is greater when all students are included in the analyses, suggesting that many of the students who had low (or negative) slopes for the Basic Skills measure subsequently dropped the class.

We were also interested in determining whether the Basic Skills and Content Analysis-Multiple Choice probes reflected growth for each type of class similarly. Table 15 reports the average slope values on each of the measures by class type.

Table 15. Descriptive Statistics for Weekly Slope Values for Both Probes by Class Type

	N	Range	Mean	SD
Basic Skills				
<i>Algebra 1A</i>	81	-.4 – 2.1	.51	.46
<i>Algebra 1B</i>	12	-.1 – 1.8	.72	.52
<i>Algebra 1</i>	12	-.8 – 1.1	.32	.51
Content Analysis-Multiple Choice				
<i>Algebra 1A</i>	80	- 1.31 – 2.26	.60	.62
<i>Algebra 1B</i>	12	-.8 – 2.25	.45	.71
<i>Algebra 1</i>	12	-.18 – 1.81	.86	.60

As Table 15 reveals, students in Algebra 1B demonstrated more growth on the Basic Skills probe than did those in the other two types of classes (.72 versus .51 and .32). However, we did not obtain a similar result on Content Analysis-Multiple Choice. On this probe, students in Algebra 1 grew more than those in other two types of classes (.86 versus .60 and .45). These two results suggest that the Basic Skills probe was more sensitive to reflecting growth for students in Algebra 1B and the Content Analysis-Multiple Choice probe was more sensitive for students in Algebra 1.

It is important to note that nearly all of the slope values for both measures either approach or exceed a weekly growth rate of .5. We have been using this benchmark as a goal in our research. We anticipate that in order for algebra progress monitoring measures to be useful to teachers on a practical level, they must be able to expect to see scores grow by at least one point every two weeks (hence a weekly growth rate of .5). The results in Table 15 are encouraging that this benchmark is achievable.

Finally, we were interested in determining whether the growth rates students obtained on the two types of probes were associated with other indicators of growth. As a result, we also examined the relationship of the slope values of both types of probes to teacher growth ratings and IAAT gain scores. These results are presented in Table 16. We found that there was a small but significant relationship between teacher growth ratings and slope values in both types of probes for all students and non-drop students. It is worth noting that non-drop students' slope values in Basic Skills had a higher correlation with teacher growth ratings than did all students (.31 versus .21 for all students). Also, table 16 reveals that slope values for both types of probes were not linearly related to IAAT gain scores. This result was surprising to us, as the IAAT had shown positive and relatively stronger relations to probe scores than other criterion measures we used. It may be that while the IAAT serves as an effective measure in dispersing students'

scores at one point in time, it is not sensitive to changes in student performance and therefore does not function well when used to create gain scores.

Table 16. Correlations Between Slope Values and Teacher Growth Ratings and IAAT Gain Scores

	All Students		Non-Drop Students	
	Teacher Growth Rating	IAAT Gain Score	Teacher Growth Rating	IAAT Gain Score
Basic Skills Slope	.21*	.07	.31**	.08
Content Analysis-Multiple Choice Slope	.22*	.03	.23*	.03

\*  $p < .05$ ; \*\*  $p < .01$

### Summary and Future Research

The main purpose of this study was to extend previous work examining the reliability and validity of the Basic Skills and the Content-Analysis-Multiple Choice probes in two Iowa schools districts and to explore the use of the measures for monitoring student progress. One hundred thirty five students in grades nine to twelve participated in the study. Data were gathered from September 2005 to January 2006. Over four months of data collection, students completed two Basic Skills probes and two Content Analysis-Multiple Choice probes each month.

We assessed the alternate form reliability and test-retest reliability of both types of probes. Our findings revealed that both types of probes possessed adequate levels of reliability, with the Basic Skills probes demonstrating a higher level of reliability than the Content Analysis-Multiple Choice probes. We specifically hypothesized that as the semester went on and students became more familiar with the probes, both types of reliability would increase. Our findings supported this hypothesis for both probes types with regard to alternate form reliability and for the Basic Skills probes with regard to test-retest reliability. However, the test-retest reliability of the Multiple Choice-Content Analysis probes (evaluated monthly) did not increase as the semester went on.

To assess the validity of both types of probes, we gathered data from a variety of indicators of students' proficiency in algebra including course grades, teachers' evaluations of student proficiency and growth, and performance on standardized assessment instruments including ITED and IAAT. We explored two types of validity: concurrent and predictive.

We assessed concurrent validity by examining the relationship of probe scores to teachers' evaluation of their students' proficiency and IAAT scores. We found that probe scores were moderately correlated with these criterion measures. When we disaggregated the data to examine concurrent validity by class type, we found that both types of probes demonstrated higher validity for Algebra 1B students than for the other two types of classes. Although the small sample sizes on which this finding is based dictate that we interpret it with caution, this finding does suggest that the criterion validity of the measures may vary for different levels of algebra and/or students with varying backgrounds in algebra.

We assessed predictive validity by investigating the relationship of the earliest probe scores (taken in the Mid-September and End-September) to teacher ratings of growth, students'

end-term algebra grades, IAAT scores taken at the end of the semester, and ITED scores. Our results revealed that in general, both probes were moderately correlated with these indicators. When examining the predictive validity of the probes using data disaggregated by class type, we again found an interesting result. Content Analysis-Multiple Choice probes demonstrated very high predictive validity for Algebra 1B students by showing a high correlation with teacher rating of growth ( $r=0.68$ ), students' end-term algebra grade ( $r=.86$ ) IAAT scores ( $r=.82$ ). As with the earlier finding, this result suggests that the probes may function differently for students of varying algebra backgrounds and/or ability levels.

We were also interested in exploring whether the Basic Skills and Content Analysis-Multiple Choice probes reflected growth similarly for each type of class. To address this issue, we conducted three sets of analyses. First, we examined graphs of the mean scores for students in each of the three class types for each data collection session. We observed a similar pattern of consistent increases on the Basic Skills measure for the Algebra 1 and Algebra 1B students. In contrast, the Algebra 1A students' means showed an initial decline, followed by small increases. On Content Analysis-Multiple Choice probes, it was found that there was an apparent difference among class types in terms of showing growth. For Algebra 1A, mean scores in the last administration were almost twice as high as the mean scores in the first administration. For the other two types of classes this ratio was about 1.5.

Next, we calculated individual student slope values for each type of probe and computed weekly rates of growth. With one exception, we found that the overall mean rates of growth, as well as those for students in each of the class types were near or above .5 units per week. The .5 threshold was not met for students in Algebra 1 on the Basic Skills measure. Finally, we were also interested in to see if the growth that students showed on both types of probes was related to their teachers' ratings of growth, and IAAT gain scores. We found that there was a small but significant correlation between teacher ratings of growth and both types of probes. No significant correlation existed between either type of probes and IAAT gain scores.

The results of this study serve to replicate previous work on the reliability and criterion validity of two of the algebra progress monitoring measures. In addition, this study moved that work forward by examining the sensitivity of the probes to changes in student performance over time and finding positive results. Future studies should replicate these findings and expand the range of probes to include the Algebra Foundations measure, which was not examined here.

# APPENDIX

Standardized Administration Directions: Basic Skills

Standardized Administration Directions: Content Analysis-Multiple Choice

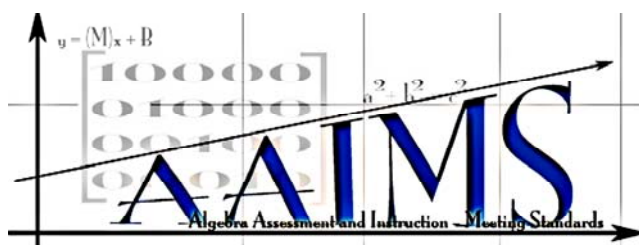
*Basic Skills* – Form 1

*Content Analysis-Multiple Choice* – Form 1

Teacher Rating of Student Proficiency

Teacher Rating of Student Growth





Project AAIMS 2005-06  
XXX High School

## PROBE STANDARD DIRECTIONS Basic Skills Probes

### GENERAL INTRODUCTION TO PROGRESS MONITORING:

The **FIRST** time you administer algebra probes, say:

*As you know, your class and other algebra classes at XXX High are working with Iowa State on a research project to learn more about improving algebra teaching and learning. Twice each month, we will be doing short algebra assessments, or probes, to monitor your learning in algebra. Remember that all students will be completing the probes and I will see the scores for all students, but your score will only be used for the Project AAIMS research if both you and your parent/guardian have given permission.*

*There are a few things you should know about these probes. First, you will be given a limited amount of time to work on the problems. These probes are different from classroom tests or quizzes and are not meant to be completely finished. What's important is that as you learn more about algebra in this class, your scores will improve. Second, keep in mind that the object of the probe is to correctly answer as many questions as you can in the amount of time given. There may be problems on the probes that are difficult or unfamiliar. Please look at each problem. If you do not know how to answer it, skip it, and go on to the next problem. DO NOT spend a great deal of time on any one problem. If you get to the end of the probe and still have time to work, go back to the problems you skipped and try to solve them. Third, your scores on these probes will be used to see your progress in algebra. Because of this, it's important that you try your best. Do you have any questions at this point?*

### BASIC SKILLS PROBES:

Hand out probe A-31 (with the sample page), keeping the probes face down. Ask students to keep the probes face down and write their name and the date on the back of the probe.

Give the standard directions:

The **FIRST** time you administer BASIC SKILLS algebra probes, say:

*Please turn your paper over. This sample page shows some examples of the types of problems on the Basic Skills probes. The questions include solving algebra equations using basic math facts, simplifying expressions by combining like terms, using the distributive property to simplify expressions, and solving proportion, or ratio problems. Now we'll take a minute so you can practice doing a Basic Skills probe. If you finish before I say 'Stop', please do NOT turn to the next page. Any questions? Ready, begin. [Time for 1 minute] Stop, pencils down.*

*Now that you've had a chance to try out this type of probe, do you have any questions? [Only answer procedural questions—do not suggest ways to solve the problems.]*

*Now we'll do the first Basic Skills probe. You will have 5 minutes to work on this two-page probe. Remember, your job is to answer as many problems correctly as you can in 5 minutes. Please look at each problem, but if you do not know how to do it, skip it and move on. If you get to the end of the probe before the time is up, go back and work on the more difficult problems. When you solve the simplifying questions, be sure to go as far as you can with your answer.*

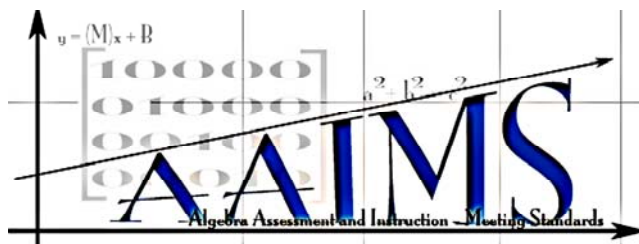
*When I say begin, please turn past the sample page and begin working. You will have 5 minutes. Please do your best work*

*Time for 5 minutes. When the timer goes off, say Stop. Please put your pencils down, and collect student papers.*

For **ALL OTHER** administrations, hand out the probes face down and say

*Please write your name and the date on the back of your paper. You are going to do a Basic Skills probe. You will have 5 minutes to work. Remember to try and complete as many problems correctly as you can in the time allowed. When you are simplifying, be sure to go as far as you can with your answer [write your answer in lowest terms]. Please do your best work. Ready? Begin.*

*Time for 5 minutes. When the timer goes off, say Stop. Please put your pencils down, and collect student papers.*



Project AAIMS 2005-06  
XXX High School

## PROBE STANDARD DIRECTIONS Textbook Probes

### E (TEXTBOOK) PROBES:

Hand out probe E-31 (with the sample page), keeping them face down. Ask students to keep the probes face down and write their name and the date on the back of the probe.

Give the standard directions:

The **FIRST** time you administer the TEXTBOOK algebra probes, say:

*The problems on this probe come from the chapters of the book, but they are not in any special order. For example, a problem from Chapter 1 could be the last problem on the probe. Please look at each problem and decide if you know how to do it. If you do, go ahead and solve the problem. If aren't certain or think you can't solve the problem, skip it and move to the next one. Don't spend too much time on any one problem. The object of the probe is to answer as many problems correctly in the time available. Once you get to the end, go back and work on the difficult problems. Remember that you may earn partial credit by showing your work even if you can't solve the entire probe. Do NOT make wild guesses because this will cause you to lose points on the probe.*

*Please turn your paper over. This sample page shows some examples of the types of problems on the Textbook probes. The problems on this probe are drawn from the different types of problems you are learning in the textbook. The questions are multiple choice. Each problem is worth 3 points, but you can earn partial credit by showing your work. Unless you are completely certain of the correct answer, the best strategy is to show your work. If you do not know the answer, you should NOT make wild guesses. You will lose points from your total score on the probe when you make wild guesses.*

*Look at the three boxes in the first row labeled A, B, and C. You'll notice that all three have answers and that the problem is the same for all three. Look at the box for Student A. She thought she knew the correct answer, so she just circled her choice at the bottom. Unfortunately, she was incorrect, so she will lose*

*a point for this problem. Student B showed his work, but did not know how to finish the problem. Because he did part of the problem correctly, Student B will earn 1 out of 3 points on this problem. Student C solved the problem, but made an error, so her final answer is not correct. Because she showed her work, she will earn 1 out of 3 points on the problem for the part she has done correctly. As you can see from these examples, it is important to show your work on these probes.*

*Let's take a minute so you can practice doing a Textbook probe. If you finish before I say 'Stop', please do NOT turn to the next page. Any questions? Ready, begin. [Time for 1 minute] Stop, pencils down.*

*Now that you've had a chance to try out this type of probe, do you have any questions? [Only answer procedural questions—do not suggest ways to solve the problems.]*

*Now we'll do the first Textbook probe. You will have 7 minutes to work on this two-page probe. Remember, your job is to answer as many problems correctly as you can in 7 minutes. Please look at each problem. If you do not know how to do it, skip it and move on. If you get to the end of the probe before the time is up, go back and work on the problems you skipped. Remember that you may earn partial credit by showing your work even if you can't solve the entire problem. Do NOT make wild guesses because this will cause you to lose points on the probe.*

*When I say begin, please turn past the sample page and begin working. You will have 7 minutes. Please do your best work. Ready? Begin.*

*Time for 7 minutes. When the timer goes off, say Stop. Please put your pencils down, and collect student papers.*

For **ALL OTHER** administrations, hand out the probes face down and say

*Please write your name and the date on the back of your paper. You are going to do a Textbook probe. You will have 7 minutes to work. Remember to try and complete as many problems correctly as you can in the time allowed. If you're not sure of an answer, skip the problem and come back to it if you have time left. Remember that you can earn partial credit by showing your work. Do NOT make wild guesses. Please do your best work. Ready? Begin.*

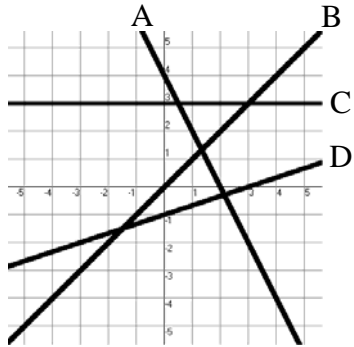
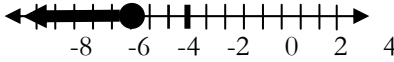
*Time for 7 minutes. When the timer goes off, say Stop. Please put your pencils down, and collect student papers.*

Solve: $9 + a = 15$ $a =$
Evaluate: $12 + (-8) + 3$
Simplify: $2x + 4 + 3x + 5$
Solve: $12 - e = 4$ $e =$
Simplify: $4(3 + s) - 7$
Simplify: $b + b + 2b$
Solve: $\frac{b}{6} = \frac{12}{18}$ $b =$
Simplify: $7 - 3(f - 2)$
Evaluate: $-5 + (-4) - 1$
Solve: $63 \div c = 9$ $c =$
Simplify: $2(s - 1) + 4 + 5s$
Simplify: $8m - 9(m + 2)$
Solve: 3 feet = 1 yard _____ feet = 9 yards
Evaluate: $4 - (-2) + 8$
Simplify: $2k + 3 - 5(k + 7)$

Solve: $10 - 6 = g$ $g =$
Simplify: $9 - 4d + 2 + 7d$
Simplify: $5(b - 3) - b$
Solve: $q \cdot 5 = 30$ $q =$
Evaluate: $8 - (-6) - 4$
Simplify: $2 + w(w - 5)$
Solve: 1 foot = 12 inches 5 feet = _____ inches
Simplify: $4 - 7b + 5(b - 1)$
Simplify: $s + 2s - 4s$
Solve: $x + 4 = 7$ $x =$
Simplify: $-5(q + 3) + 9$
Evaluate: $9 + (-3) - 8$
Solve: $\frac{12}{2} = \frac{48}{e}$ $e =$
Simplify: $y^2 + y - 4y + 3y^2$
Simplify: $3(c + 2) - 2c$

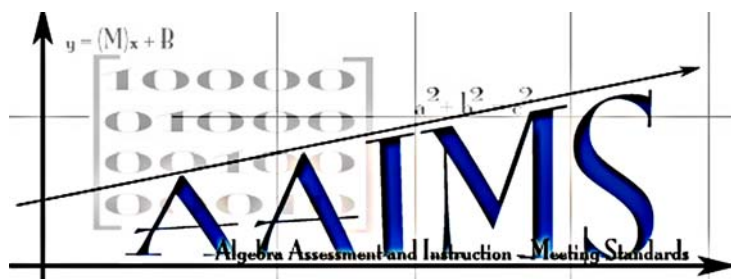
Solve: $3 \cdot 8 = m$ $m =$
Evaluate: $-9 + 5 + 8$
Simplify: $x + 2(x - 5) - 3$
Solve: $d - 5 = 4$ $d =$
Simplify: $5(3 + f) - 2f + 6$
Simplify: $5 - 2b + 4(b + 3)$
Solve: 4 quart = 1 gallon _____ quarts = 3 $\frac{1}{4}$ gallons
Simplify: $4(y + 1) - 8y$
Evaluate: $14 - 7 + (-3)$
Solve: $\frac{36}{6} = s$ $s =$
Simplify: $-3w^2 + 5w^2 - 5 + 12$
Simplify: $9 - 4(v + 2)$
Solve: $4r = 28$ $r =$
Simplify: $16 + 2(t - 4) - 3t$
Simplify: $c - 3(c + 2) + 8$

Solve: $\frac{1.5}{3} = \frac{h}{9}$ $h =$
Simplify: $7b - 4 - 3 - 2b$
Simplify: $2e - 3(e - 4)$
Solve: $6 + 7 = v$ $v =$
Evaluate: $-5 + 6 - 6$
Simplify: $4 + 10(1 - r)$
Solve: $2.5 \text{ cm} = 1 \text{ inch}$ _____ cm = 6 inches
Simplify: $6a + 2a - 9 + 3a^2$
Evaluate: $-1 + 4 + (-7)$
Solve: $\frac{500}{j} = \frac{10}{2}$ $j =$
Simplify: $-3(u + 3) - 2u + 5$
Simplify: $2c - 3c - c$
Solve: $h \div 6 = 8$ $h =$
Evaluate: $-2 + (-5) + (-8)$
Simplify: $3z - 8z + 2 + 9$

<p>Solve: <math>3x + 4 = 19</math></p> <p>a) <math>x = 8</math> b) <math>x = 22</math> c) <math>x = 15</math> d) <math>x = 5</math></p>	<p>Evaluate <math>a^2 - b \div 2</math> when <math>a = 4</math> and <math>b = 6</math></p> <p>a) 1 b) 5 c) 10 d) 13</p>	<p>Which line on the graph is <math>y + 2x = 4</math> ?</p>  <p>a) Line A b) Line B c) Line C d) Line D</p>	<p>Simplify: <math>3(m + 2) + 2(m - 1)</math></p> <p>a) <math>5m + 4</math> b) <math>5m + 1</math> c) <math>6m + 8</math> d) <math>6m - 8</math></p>
<p>Evaluate the expression: <math>4^{-2}</math></p> <p>a) <math>-16</math> b) <math>\frac{1}{16}</math> c) <math>\frac{1}{8}</math> d) <math>-8</math></p>	<p>Solve the linear system: <math>x - y = 4</math> <math>x + 2y = 19</math></p> <p>a) <math>(-1, -5)</math> b) <math>(5, 8)</math> c) <math>(-2, 19)</math> d) <math>(9, 5)</math></p>	<p>This graph shows the solution for which equation?</p>  <p>a) <math>x &gt; -3</math> b) <math>2x \leq -6</math> c) <math>-3x &gt; 9</math> d) <math>3x \geq 9</math></p>	<p>Write the equation in slope-intercept form if <math>m = \frac{1}{2}</math> and <math>b = 3</math></p> <p>a) <math>y = 2x + 3</math> b) <math>y = 3x + \frac{1}{2}</math> c) <math>x = \frac{1}{2}y - 3</math> d) <math>y = \frac{1}{2}x + 3</math></p>

<p>Evaluate <math>d + 3e^2</math> when <math>d = 5</math> and <math>e = 2</math></p> <p>a) 11 b) 23 c) 17 d) 10</p>	<p>Solve: <math>6c + 4 = -3c - 14</math></p> <p>a) <math>-\frac{10}{3}</math> b) <math>-2</math> c) 2 d) 6</p>	<p>Find the slope of a line through <math>(1, -1)</math> <math>(5, 2)</math></p> <p>a) <math>\frac{1}{5}</math>      b) <math>\frac{3}{4}</math> c) <math>-6</math>      d) <math>-\frac{4}{3}</math></p>	<p>Simplify: <math>6(2b - 3) - 3(2 - b)</math></p> <p>a) <math>15b - 24</math> b) <math>9b - 9</math> c) <math>9b + 12</math> d) <math>15b + 12</math></p>
<p>Simplify the expression:</p> $\frac{a^2}{ab^3} \cdot \frac{b^4}{a^3}$ <p>a) <math>\frac{a^8}{a^3b^3}</math>      b) <math>\frac{ab^8}{a^4b^3}</math> c) <math>\frac{b}{a^2}</math>      d) <math>\frac{b}{a}</math></p>	<p>Solve the linear system: <math>-6x + 3y = -6</math> <math>2x + 6y = 30</math></p> <p>a) (6, 3) b) (3, 4) c) (2, 6) d) (4, -3)</p>	<p>Simplify <math>b^2 - 4b + 2b^2 + 7 - 5</math></p> <p>a) <math>3b^2 - 4b + 2</math> b) <math>2b + 2</math> c) <math>-b^2 - 4b + 12</math> d) <math>3b^2 - 4b + 12</math></p>	<p>Write the equation of a line through <math>(5, 3)</math> <math>(4, 9)</math>. Use point-slope form.</p> <p>a) <math>y + 1 = 2(x - 4)</math> b) <math>y + 4 = -6(x - 1)</math> c) <math>y - 3 = -6(x - 5)</math> d) <math>y = -6x + 30</math></p>





Teacher \_\_\_\_\_

**Project AAIMS:**  
**Algebra Assessment and Instruction:**  
**Meeting Standards**  
 XXX High School  
 Fall 2005

Directions: For each student, rate his or her **general proficiency in algebra** relative to other students in your class(es). A rating of "1" indicates a very low level of proficiency, "4" indicates average proficiency, and "7" indicates exceptional proficiency. Try to spread student ratings across the **full range of the scale**, not clustering students only in the middle or toward one end.

Student

Algebra Proficiency

Low		Average				High	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	
1	2	3	4	5	6	7	

Teacher \_\_\_\_\_

**Project AAIMS:**  
**Algebra Assessment and Instruction:**  
**Meeting Standards**  
XXX Senior High

## Teacher Rating of Student Progress

**Directions:** Below is a list of the students you teach. Please rate the amount of progress or improvement each student has made in algebra during this course. A rating of "1" indicates no growth or a decrease in level of performance, "3" indicates average progress, and "5" indicates exceptional progress, far beyond what you expected. If the list includes a student who has dropped the class, just draw a line through his/her name. Thank you!

<u>Student</u>	<u>Algebra Progress</u>				
	Low	Average			High
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5